

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
7 April 2005 (07.04.2005)

PCT

(10) International Publication Number
WO 2005/031600 A2

(51) International Patent Classification⁷: **G06F 17/30**

[RU/GB]; 3 Railway Cottages, Greenisland, Carrickfergus,
County Antrim BT38 8RB (GB).

(21) International Application Number:

PCT/EP2004/010877

(74) Agents: WALLACE, Alan et al.; F.R. Kelly & Co., 4
Mount Charles, Belfast BT7 1NZ (GB).

(22) International Filing Date:

27 September 2004 (27.09.2004)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

0322600.8 26 September 2003 (26.09.2003) GB

(71) Applicants (for all designated States except US): UNI-
VERSITY OF ULSTER [GB/GB]; Cromore Road, Col-
eraine, County Londonderry BT52 1SA (GB). ST. PE-
TERSBERG STATE UNIVERSITY [RU/RU]; Univer-
sitetskii Prospekt, 28, Petergof, St. Petersburg, 198504
(RU).

(72) Inventors; and

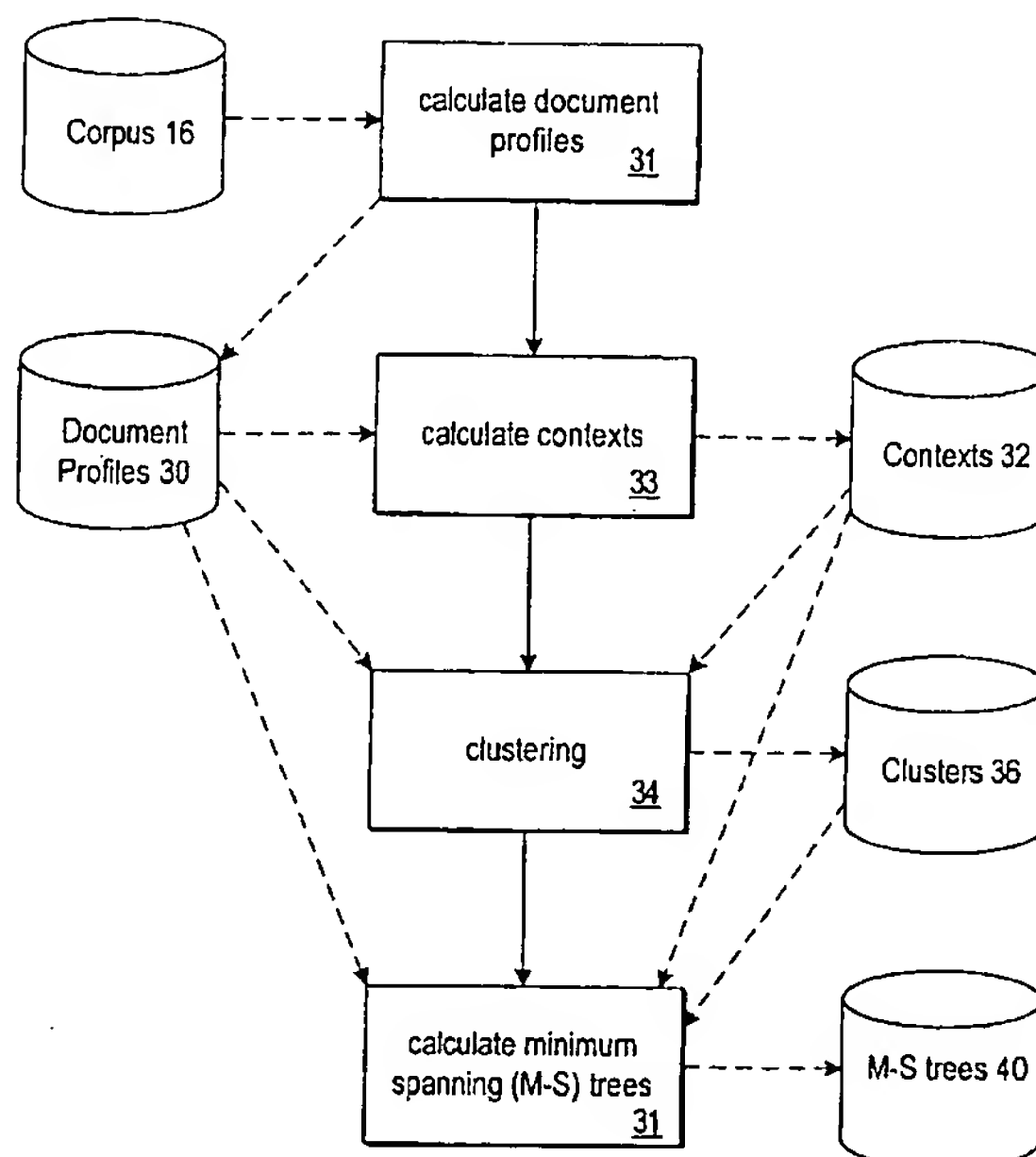
(75) Inventors/Applicants (for US only): PATTERSON,
David [GB/GB]; 7 Wood Grange, Jordanstown, County
Antrim BT37 0WG (GB). DOBRYNIN, Vladimir

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM,
ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI,
SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: COMPUTER AIDED DOCUMENT RETRIEVAL



(57) Abstract: A method of determining cluster attractors for a plurality of documents comprising at least one term. The method comprises calculating, in respect of each term, a probability distribution indicative of the frequency of occurrence of the, or each, other term that co-occurs with said term in at least one of said documents. Then, the entropy of the respective probability distribution is calculated. Finally, at least one of said probability distributions is selected as a cluster attractor depending on the respective entropy value. The method facilitates very small clusters to be formed enabling more focused retrieval during a document search.

WO 2005/031600 A2

**Declaration under Rule 4.17:**

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE,*

BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

- *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.